

This Raw Listing contains the General  
Information Section and up to the first 5 pages.

MAY 01 2000  
TECH CENTER 1600/2901

## SEQUENCE LISTING

ENTERED

1  
2 (1) General Information:  
3  
4 (i) APPLICANT: Choulika, Andre  
5 Perrin, Arnaud  
6 Dujon, Bernard  
7 Nicolas, Jean-Francois  
8  
9 (ii) TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme  
10 I-SCEI and the Uses Thereof  
11  
12 (iii) NUMBER OF SEQUENCES: 52  
13  
14 (iv) CORRESPONDENCE ADDRESS:  
15 (A) ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
16 Dunner  
17 (B) STREET: 1300 I Street, N.W.  
18 (C) CITY: Washington  
19 (D) STATE: D.C.  
20 (E) COUNTRY: USA  
21 (F) ZIP: 20005-3315  
22  
23 (v) COMPUTER READABLE FORM:  
24 (A) MEDIUM TYPE: Floppy disk  
25 (B) COMPUTER: IBM PC compatible  
26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25  
28  
29 (vi) CURRENT APPLICATION DATA:  
30 (A) APPLICATION NUMBER: Unassigned  
31 (B) FILING DATE: 25-JAN-2000  
32 (C) CLASSIFICATION:  
33  
34 (vii) PRIOR APPLICATION DATA:  
35 (A) APPLICATION NUMBER: US 09/244,130  
36 (B) FILING DATE: 04-FEB-1999  
37  
38 (vii) PRIOR APPLICATION DATA:  
39 (A) APPLICATION NUMBER: US 09/119,024  
40 (B) FILING DATE: 20-JUL-1998  
41  
42 (vii) PRIOR APPLICATION DATA:  
43 (A) APPLICATION NUMBER: US 08/336,241  
44 (B) FILING DATE: 07-NOV-1994  
45  
46

--&gt;

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/492,697

MAY 01 2000  
TECH CENTER 1600/2900  
DATE: 02/16/2000  
TIME: 02:14:57

INPUT SET: S34756.raw

47 (vii) PRIOR APPLICATION DATA:  
48 (A) APPLICATION NUMBER: US 07/971,160  
49 (B) FILING DATE: 05-NOV-1992

50 (vii) PRIOR APPLICATION DATA:  
51 (A) APPLICATION NUMBER: US 07/879,689  
52 (B) FILING DATE: 05-MAY-1992

53 (viii) ATTORNEY/AGENT INFORMATION:  
54 (A) NAME: Meyers, Kenneth J.  
55 (B) REGISTRATION NUMBER: 25,146  
56 (C) REFERENCE/DOCKET NUMBER: 3495-0111-11

57 (ix) TELECOMMUNICATION INFORMATION:  
58 (A) TELEPHONE: 202-408-4000  
59 (B) TELEFAX: 202-408-4400

60 (2) INFORMATION FOR SEQ ID NO:1:  
61  
62  
63  
64 (i) SEQUENCE CHARACTERISTICS:  
65 (A) LENGTH: 714 base pairs  
66 (B) TYPE: nucleic acid  
67 (C) STRANDEDNESS: single  
68 (D) TOPOLOGY: linear  
69  
70 (ii) MOLECULE TYPE: DNA (genomic)  
71  
72  
73 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
74  
75 ATGCATATGA AAAACATCAA AAAAAACCAAG GTAATGAACC TCGGTCCGAA CTCTAAACTG 60  
76 CTGAAAGAAT ACAAAATCCCA GCTGATCGAA CTGAACATCG AACAGTCGA AGCAGGTATC 120  
77 GGTCTGATCC TGGGTGATGC TTACATCCGT TCTCGTGATG AAGGTAAAAC CTACTGTATG 180  
78 CAGTCGAGT GGAAAAACAA AGCATAACATG GACCACGTAT GTCTGCTGTA CGATCAGTGG 240  
79 GTACTGTCCC CGCCGCACAA AAAAGAACGT GTTAACCACC TGGGTAACCT GGTAATCACC 300  
80 TGGGGCGCCC AGACTTTCAA ACACCAAGCT TTCAACAAAC TGGCTAACCT GTTCATCGTT 360  
81 AACAAACAAA AAACCATCCC GAACAACCTG GTTGAAAAC ACCTGACCCC GATGTCTCTG 420  
82 GCATACTGGT TCATGGATGA TGGTGGTAAA TGGGATTACA ACAAAAACTC TACCAACAAA 480  
83 TCGATCGTAC TGAACACCCA GTCTTCACT TTCAAGAAG TAGAATACCT GTTTAAGGGT 540  
84 CTGCGTAACA AATTCCAAC TGAACGTTAC GTAAAAATCA ACAAAAACAA ACCGATCATC 600  
85 TACATCGATT CTATGTCTTA CCTGATCTTC TACAACCTGA TCAAACCGTA CCTGATCCCG 660

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/492,697DATE: 02/16/2000  
TIME: 02:14:57

INPUT SET: S34756.raw

100 CAGATGATGT ACAAACTGCC GAACACTATC TCCTCCGAAA CTTTCCTGAA ATAA 714  
101  
102  
103 (2) INFORMATION FOR SEQ ID NO:2:  
104 (i) SEQUENCE CHARACTERISTICS:  
105 (A) LENGTH: 237 amino acids  
106 (B) TYPE: amino acid  
107 (D) TOPOLOGY: linear  
108  
109 (ii) MOLECULE TYPE: peptide  
110  
111  
112  
113 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:  
114 Met His Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro  
115 15  
116 5  
117 1  
118 Asn Ser Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn  
119 30  
120 20 25  
121 Ile Glu Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr  
122 45  
123 35  
124 Ile Arg Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp  
125 60  
126 50 55  
127 Lys Asn Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp  
128 80  
129 65 70 75  
130 Val Leu Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn  
131 95  
132 85 90  
133 Leu Val Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn  
134 110  
135 100 105  
136 Lys Leu Ala Asn Leu Phe Ile Val Asn Asn Lys Thr Ile Pro Asn  
137 125  
138 115 120  
139 Asn Leu Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe  
140 140  
141 130 135  
142 Met Asp Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys  
143 160  
144 145 150 155  
145 Ser Ile Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Val Glu Tyr  
146 175  
147 165 170  
148 Leu Val Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys  
149 190  
150 180 185  
151 Ile Asn Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu  
152

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/492,697DATE: 02/16/2000  
TIME: 02:14:58

INPUT SET: S34756.raw

153	195	200	205
154	Ile Phe Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr		
155	210 215 220		
156			
157	Lys Leu Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys		
158	225 230 235		
159			

160 (2) INFORMATION FOR SEQ ID NO:3:

161 (i) SEQUENCE CHARACTERISTICS:  
 162 (A) LENGTH: 722 base pairs  
 163 (B) TYPE: nucleic acid  
 164 (C) STRANDEDNESS: single  
 165 (D) TOPOLOGY: linear  
 166  
 167 (ii) MOLECULE TYPE: DNA (genomic)

168 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

173	AAAAATAAAA TCATATGAAA AATATTAAA AAAATCAAGT AATCAATCTC GGTCTTATT	60
174	CTAAATTATT AAAAGAATAT AAATCACAAT TAATTGAATT AAATATTGAA CAATTGAAAG	120
175	CAGGTATTGG TTTAATTGTTA GGAGATGCTT ATATTCGTAG TCGTGATGAA GGTAAAAC	180
176	ATTGTATGCA ATTTGAGTGG AAAAATAAGG CATAATGGA TCATGTATGT TTATTATATG	240
177	ATCAATGGGT ATTATCACCT CCTCATAAA AAGAAAGAGT TAATCATTAA GGTAATTTAG	300
178	TAATTACCTG GGGAGCTCAA ACTTTAAC ATCAAGCTTT TAATAAATTA GCTAACTTAT	360
179	TTATTGTAAA TAATAAAAAA CTTATTCTA ATAATTAGT TGAAAATTAT TTAACACCTA	420
180	TGAGTCTGGC ATATTGGTT ATGGATGATG GAGGTAAATG GGATTATAAT AAAAATTCTC	480
181	TTAATAAAAG TATTGTATTA AATACACAAA GTTTACTTT TGAAGAAGTA GAATATTTAC	540
182	TTAAAGGTTT AAGAAATAAA TTTCAATTAA ATTGTTATGT TAAAATTAAT AAAAATAAAC	600
183	CAATTATTAA TATTGATTCT ATGAGTTATC TGATTTTTA TAATTTAATT AACCTTATT	660
184	TAATTCCCTCA AATGATGTAT AAACTGCCTA ATACTATTC ATCCGAAACT TTTTAAAT	720
185		722

186 AA

187 (2) INFORMATION FOR SEQ ID NO:4:

188 (i) SEQUENCE CHARACTERISTICS:  
 189 (A) LENGTH: 235 amino acids  
 190 (B) TYPE: amino acid

RAW SEQUENCE LISTING  
PATENT APPLICATION US/09/492,697DATE: 02/16/2000  
TIME: 02:14:58

INPUT SET: S34756.raw

206 (D) TOPOLOGY: linear  
207  
208 (ii) MOLECULE TYPE: peptide  
209  
210  
211 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:  
212  
213 Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser  
214 1 5 10 15  
215  
216 Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu  
217 20 25 30  
218  
219 Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg  
220 35 40 45  
221  
222 Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn  
223 50 55 60  
224  
225 Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu  
226 65 70 75 80  
227  
228 Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val  
229 85 90 95  
230  
231 Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu  
232 100 105 110  
233  
234 Ala Asn Leu Phe Ile Val Asn Asn Lys Leu Ile Pro Asn Asn Leu  
235 115 120 125  
236  
237 Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp  
238 130 135 140  
239  
240 Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Leu Asn Lys Ser Ile  
241 145 150 155 160  
242  
243 Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Cys Tyr Leu Val  
244 165 170 175  
245  
246 Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys Ile Asn  
247 180 185 190  
248  
249 Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe  
250

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/09/492,697**

DATE: 02/16/2000  
TIME: 02:14:58

*INPUT SET: S34756.raw*

Line	Error	Original Text
31	Wrong application Serial Number	(A) APPLICATION NUMBER: Unassigned